30/11/2025

Iterative Relevance Feedback in LifeLog Search Utilising EEG

Aaron Gillespie - aaron.gillespie9@mail.dcu.ie
Shahin de Faoite Houshidari - shahin.defaoitehoushidari2@mail.dcu.ie
Sadhbh Murphy - sadhbh.murphy393@mail.dcu.ie
Ciaran Farrell - ciaran.farrell94@mail.dcu.ie

Iterative Relevance Feedback in LifeLog Search Utilising EEG © 2025 by Sháhín de Faoite Houshidari, Ciarán Farrell, Sadhbh Murphy, Aaron Gillespie is licensed under CC BY-NC-SA 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-sa/4.0/

Table of Content

Table of Content **2**
Editors Note: 4
Glossary of Terms 5
Abstract 7
1. Introduction 8
1.1. Related Work 8
1.1.1 RF, Multimedia IR 8
1.1.2 Lifelogging 8
1.1.3. BCIs 9
2. User Analysis 10
2.1 Target User 10
2.2. User Needs 11
2.3 Addressing User Needs 12
3. Scientific Functional Description 14
3.1 Overview 14
3.1.1. Data Flow 17
3.1.1.1. Initial Ranking 18
3.1.1.2. IRF RSVP Layer 18
3.1.1.3. IRF Results Layer 19
3.1.1.4. Generating Binary Scores using Brain-Signals 19
3.2 System Architecture 21
3.2.1. System Architecture Diagram 21
3.2.2. Component Description 21
3.2.2.1. ReactJS UI 21
3.2.2.1.1. Query Input 21
3.2.2.1.2. Signal Collection 21
3.2.2.1.2.1. Layer 1 (RSVP) 21
3.2.2.1.2.1. Layer 2 (BCI, Top-K, Click) 22
3.2.2.2. Signal Processing 22
3.2.2.3. Fusion 22
3.2.2.4. Query Reformulation 22
3.2.2.5. Query Parser 22
3.2.2.6. Results Merger 22
3.3. Algorithms 23
3.3.1. Signal Fusion 23
3.4. Design Analysis and Justification 24
3.4.1. CLIP 24
3.4.2. Multi-Layer RF using BCIs 24
3.4.3. Cosine Similarity 24
3.4.4. EEG and Laboratory Environment 25
3.4.5. Milvus 25
3.4.6. User-Interface 25
3.4.7. SVM V.S. CNN for BCI Model 25
3.4.8. SVM Kernel 25
3.4.9. Query Parsing and Interleaving 26
4. Evaluation Plan 26

4.1 Evaluation Design.....26
4.1.1. A/B/n Testing.....27
4.1.2. Usability Testing.....27
4.1.3. A/B Testing.....27
4.1.4. Hyper-parameter evaluation.....28
4.2 Data and Collection.....28
4.2.1. LifeLog Dataset.....28
4.2.2. Collecting Signals.....29
4.2.2.1. Brain-Signals.....29
4.2.2.2. Click-Signals and Top-K.....30
4.3 Criteria and metrics.....31
4.3.1. Quantitative.....31
4.3.2. Qualitative.....31
5. Future Work.....31
6. Conclusion.....32
7. Group Activity Summary.....33
8. References.....35
9. Bibliography.....39
10. Appendix.....39
10.1. Heuristic Analysis.....40
10.2. Example Queries for System.....41
10.3. SUS Survey with Subjective Feedback Questions.....42

Preface:

This paper was prepared by four undergraduate Computer Science students studying in Dublin City University as part of their coursework for CSC1103, “Search Technologies”. It is shared publicly to encourage further research in utilising Brain-Computer-Interfaces to tackle problems in Information Retrieval.

This work is protected under the Creative Commons CC BY-NC-SA 4.0 License. This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only. If others modify or adapt the material, they must license the modified material under identical terms.

Glossary of Terms

Term	Definition
IR	Information Retrieval
BCI	Brain-Computer Interface
EEG	Electroencephalogram: a type of Brain Computer Interface to record the electrical activity of the brain, using electrodes on brain-caps to collect signals.
ASK	“Anomalous State of Knowledge”
RF	Relevance Feedback: a method of local query reformulation where subsequent results are affected by the manner in which a User interacts with the first set of results.
IRF	Iterative Relevance Feedback (RF that occurs iteratively during the session).
RRF	Retrospective Relevance Feedback (RF that occurs subsequent to termination of the session, wherein results affect future sessions).
KIS	Known-Item Search
AD	Ad-Hoc Search
QA	Query-Answering
RAG	Retrieval Augmented Generation
LSC’25	Lifelog Search Challenge 2025
SVM	State Vector Machine
CNN	Convolutional Neural Network
SSVEP	Steady-State Visually Evoked Potential. SSVEPs are the neural responses to visual stimulation at specific frequencies.
NFR	Non-Functional-Requirements

FE	Feature-Extraction. As it pertains to BCIs, this involves extracting relevant “features” from an EEG scan.
DE	Differential Entropy
RSVP	Rapid-Serial-Visual-Presentation: wherein images are shown *rapidly* to a User to collect brain-signals related to the image shown. This allows for *search* and *selection* at rapid speed by analyzing brain-responses to the visual stimuli.

Abstract

LifeLogging involves the capturing of “memory-images” from a first-person-perspective using a wearable camera to develop a searchable “memory-archive”, where offsetting “ambiguous queries” remains problematic. A User may have *some* information in relation to the “memory” they are trying to fetch, but not all; especially with compensating for “memory difficulties” being a primary use-case for LifeLogging. This naturally leads to “query ambiguity”.

Relevance Feedback (RF) is typically deployed to offset this, but relies either on explicit User Feedback (for which a User is *consciously* responsible for, increasing cognitive load), or “implied relevance” gauged via click-feedback (which can be misleading).

Recent studies in Brain-Computer-Interfaces have demonstrated the utility of EEG in recognising and predicting relevance through the processing of Event-Related-Potential (ERP) effects; signals that are *authentic* responses to visual stimuli and how they relate to their information requirement.

We propose the BCI-RF-LifeLogging system, a multimedia retrieval system that utilises BCIs to perform a natural derivation of the ASK and efficiently offset query ambiguity. Considering the problem of LifeLogging (wherein a User has *recognition* if not *perfect recall),* this system is a natural fit.

We introduce a novel, multi-layered approach, building on recent studies in BCIs and LifeLogging, wherein a User engages in *one layer* of Rapid-Serial-Visual-Presentation (RSVP) to refine the candidate list, and a second layer of *exploratory-search* wherein brain-signals, click-signals, and pseudo-relevant signals (top-k) are collected and fused to generate relevance scores. *Features* (objects, scenes, time) related to relevant images are subsequently used for query-expansion.

1. Introduction

The problem of the “ambiguous query” is prevalent in Information Retrieval (IR), particularly in multimedia search contexts. Consider LifeLogging (capturing “memory-images” with a wearable camera, saved to a searchable memory-database), where Users are familiar with the content in the dataset without necessarily being able to perfectly describe items: Users must engage in Known-Item-Search (KIS) and time-consuming exploratory-search in order to concretize their Anomalous-State-Of-Knowledge (ASK). Relevance Feedback (RF) can be used to re-rank results to combat ambiguity, but this is reliant on the User supplying difficult-to-collect [1] and time-consuming explicit-signals or potentially misleading implicit-signals [2].

We introduce a novel RF framework for LifeLog search, utilising brain-signals to enrich implicit-signals (such as “click”), sidestep explicit-feedback in exploratory-search, speed up KIS, and combat query-ambiguity, wherein exploration is guided by brain-signals combined with pseudo-relevant and click-signals.

We use LifeLogging as a prescient case-study with considerable motivation - “replaying” a day’s activities has been found to improve memory recall (even in those with moderate Alzheimer's disease) [3]. In preparing this study, we acknowledge the constraints inherent in EEG-based studies; principally, the variability of brain-signals. As such, we restrict our evaluation to lab-based settings.

1.1. Related Work

1.1.1 RF, Multimedia IR

Nara et al. [4] demonstrate the viability of utilising RF in the context of image-search utilising CLIP. They note RF is particularly useful in modelling User Preference. Their system fetches images, receives explicit binary feedback from the User, and learns preference.

Content-Based-Image-Retrieval (CBIR) matches queries to images based on the *content* of an image [5], extracting context to supplement searchable metadata.

Models like CLIP [6] are embedding-based systems that map text and images into a shared embedding space, permitting the calculation of cosine-similarity between query input-vectors, and image-vectors.

1.1.2 Lifelogging

The annual Lifelog Search Challenge (LSC) [7] is a conference for developing Lifelog-search solutions. At LSC’25, the MemoriEase system [8] boasted impressive results in QA using RAG, while utilizing traditional RF for KIS and AD tasks.

SnapSeek 3.0 [9] reported gains over MemoriEase 3.0 in KIS and Ad-Hoc-Search (AD) through extracting features (labelling activities, time of day, word frequency analysis, representing

image-objects with proper-nouns). The Memoria system [10] (which received the highest score in AD tasks in LSC'25) generated textual-annotations using object-detection, text-recognition, image-captioning, scene-recognition and temporal information. This information was stored in PostgreSQL.

Additionally, Memoria utilised an event-segmentation structure, clustering images into blocks called "events" (also stored in PostgreSQL). This improves AD by allowing the User to search for events. MemoriEase *removed* event grouping, which is relevant when considering the disparity in performance for the AD task between MemoriEase and Memoria.

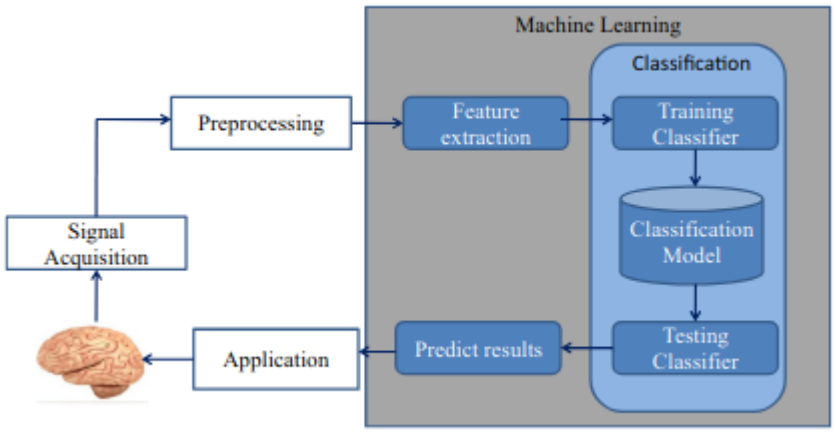
Lumina-1 (also of LSC'25) re-ranks search results based on explicit user-defined relevance [11], demonstrating the potential of RF in LifeLogging.

Memoria [10] and MemoriEase [8] utilise CLIP, where Lumina-1 [11] and SnapSeek 3.0 [9] utilise CLIPS (a CLIP framework for learning with synthetic captions). All three utilise Milvus vector databases.

At query time, Memoria [10] used *DeepSeek* to extract the "main-topic" of some query, and "*Stanza"* to extract relevant "entities" (such as time, objects, or scenes). Likely relevant candidates were fetched from Milvus by utilising the cosine similarity of the CLIP embedding of the *"main topic",* where these results were interleaved with candidates fetched from PostgreSQL based on the "entities" extracted by *Stanza.* This was done to alleviate the inherent loss in detail when embedding images.

1.1.3. BCIs

Brain-Computer-Interfaces (BCIs) digitally record neural-activity, and have been deployed in a variety of contexts including IR [12] (specifically RF, where signals are passed to a Machine-Learning model that matches a User's brain-signal-characteristics to labeled data [2]).



Data-flow diagram for brain-signal processing, Aggarwal et al. [31]

As noted by Aggarwal et al. [13], The ML-BCI data-flow involves (sequentially):

1. Acquiring signals

    a. Utilising brain-caps hooked up to NuAmps systems [2].
2. Preprocessing
    a. Denoising brain-signals to reduce variability.
3. Feature extraction
    a. Generating vectors representing relevant features.
4. Classification
    a. Generating indicators of relevance.

BCIs are uniquely suited to automatic-labelling tasks: the brain automatically generates ERP effects, such as the P300 (which registers 300ms after noticing something relevant [14]) which can be correlated to relevance. Ye et al. [2] propose an IR system combining BCI-based RF with pseudo-relevance and implicit-signals to improve document re-ranking in text-based-retrieval. BCI-signals improved RF across several search scenarios - particularly in contexts where pseudo-relevant or implicit-signals might mislead.

Moreover, Eugster et al. [15] demonstrate utilising brain-signals to infer document relevance utilising Wikipedia articles. This study adds that brain-interfaces sidestep the requirement of utilising *explicit user-interaction* to gauge relevance.

Additionally, Quach et al. demonstrate the viability of utilising BCI-signals for "predicting" relevant words through their DERCo dataset [16], with Quach outlining a system to predict *next-word-relevance* in a manner similar to an LLM utilising EEG-signals [17].

Finally, Healy et al. [18] demonstrate utilising BCI-signals to sort life-log retrieval results, using Rapid-State Visually-evoked Phenomenon (RSVP) (wherein visual stimuli are flashed rapidly, and the brain generates a response that represents the relevance or the recognition of some visual phenomenon) to extract brain-responses.

2. User Analysis

2.1 Target User

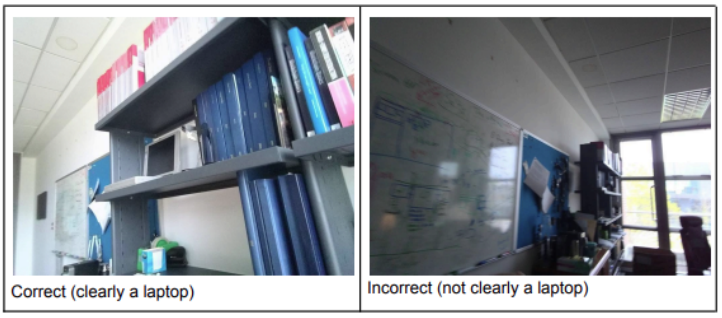
Our target user is a *LifeLogger* who uses a retrieval system to extract images from a dataset of "memory images", where data is recorded from the user's point-of-view.

It is assumed that the user has technical proficiency equivalent to that of an average university student.

We assume the user is a search novice (compared to a librarian). We do not assume the user is capable of preparing precise queries, or possesses perfect recall.

Considering the user of a LifeLogging system (and that people with *memory difficulty* would especially find such systems useful) our problem is relevant to our Target User.

LSC24-AD06
I used to have an old grey laptop (from the 1990s). It used to be on a shelf in my office. Can you find it? The laptop must be visible and clearly a laptop.



Correct (clearly a laptop) | Incorrect (not clearly a laptop)

Figure 1: LSC'24 AD06 [19]

LSC24-AD04
I can not find my electronic violin. I left it with my cream coloured electric guitar (Telecaster style). Can you find images of them together?

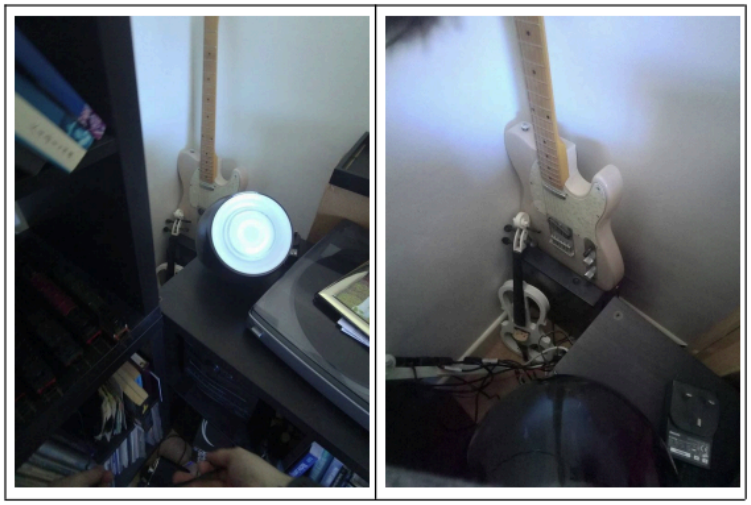


Figure 2: LSC'24 AD04 [19]

Figures 1 and 2: AD search tasks that illustrate our target User's use-case for this system. They may have limited information, so the system must handle ambiguity.

The User is medically cleared for EEG usage [20]. Our target User speaks English.

2.2. User Needs

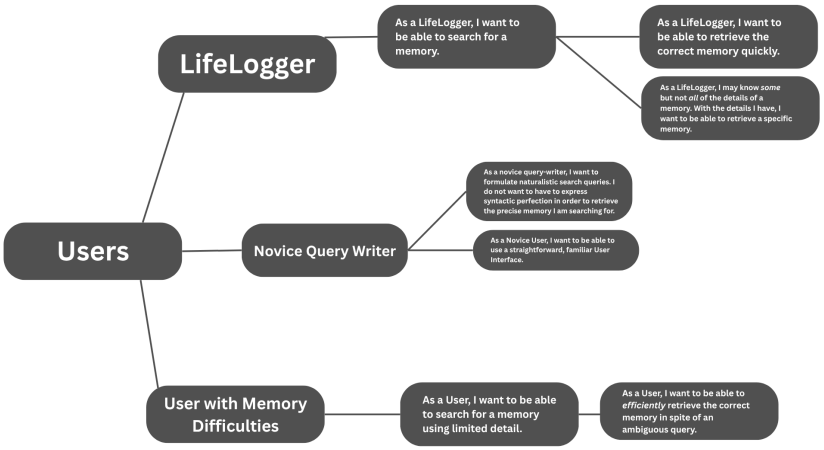


Figure 3. User Stories Diagram

Users require:

- Simple interface for efficient and accurate retrieval.
- Offsetting of query ambiguity.
    - Users may not possess strong recall (that may be why they are a LifeLogger).
- Filters to refine search
    - If the User, while exploring, recalls details that they lacked awareness of when initially performing query formulation (date or time), they must be allowed to "filter" results.

2.3 Addressing User Needs

We direct stakeholders to *Section 3.1* for User-Interface mockups. We address *User needs* through *Iterative-Relevance-Feedback* (IRF) (fusing click-signals, pseudo-relevance, and EEG-signals to de-emphasise precise query formulation and reduce cognitive load), a *simple* User-Interface, multimedia search (a User can search using *text* for an *image*), and letting users filter results by date or time.

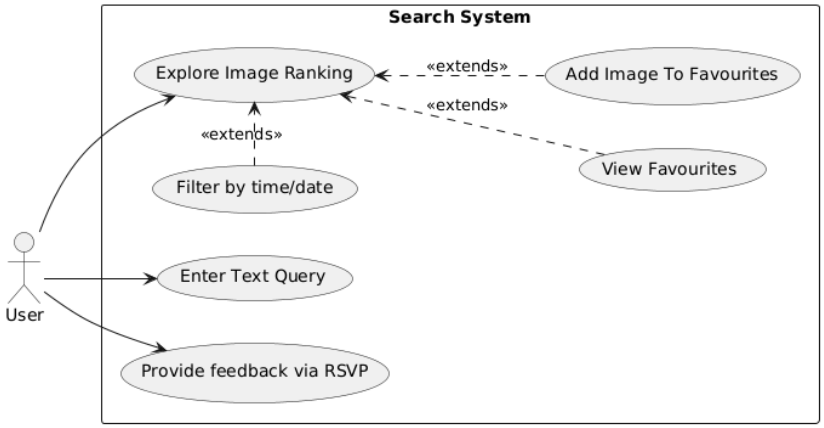


Figure 4. Use Case Diagram

3. Scientific Functional Description

3.1 Overview

We broadly overview the system, using User-Interface wireframes to illustrate. Afterwards, we justify design decisions, describe data-flow, and outline algorithms.

1. User Submits Query

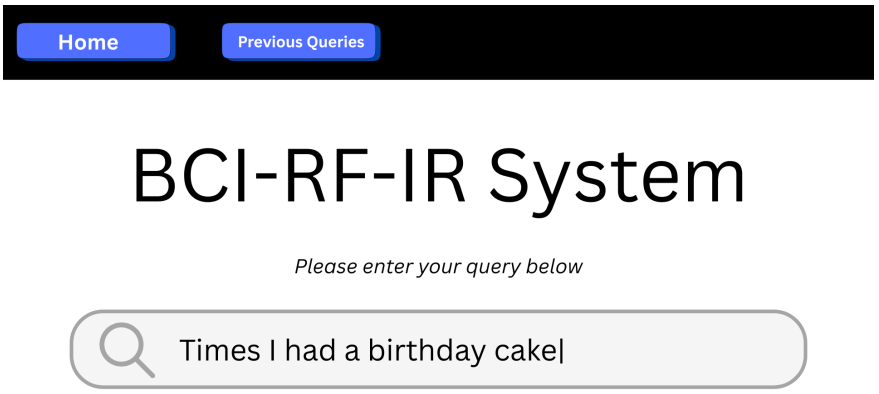


Figure 5: Home

2. User engages in RSVP

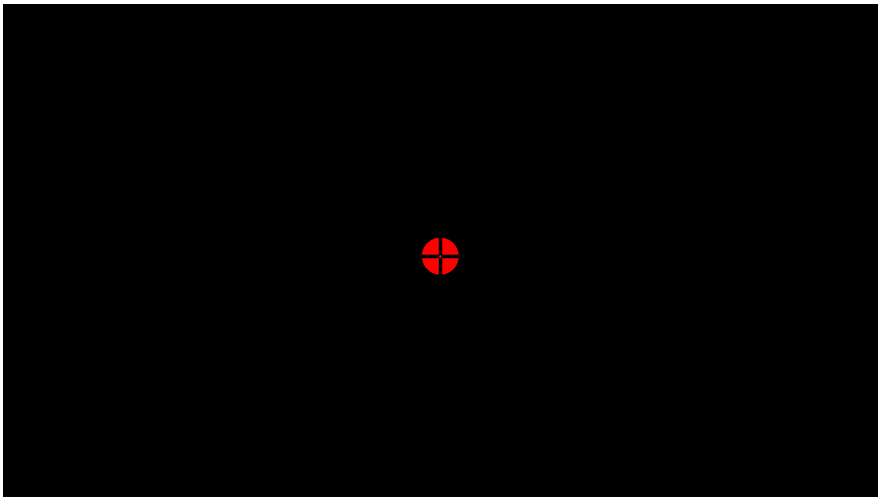


Figure 6.1: "Red Fixation Cross" indicating where the image shall appear [21].

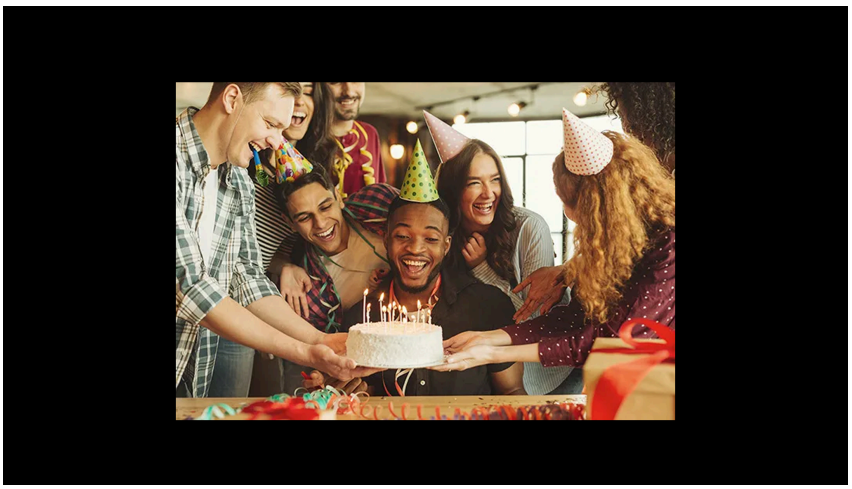


Figure 6.2

Images are rapidly shown to the user (RSVP) and brain-signals are collected such that relevance can be ascertained from brain-responses to visual-stimuli.

3. User browses results

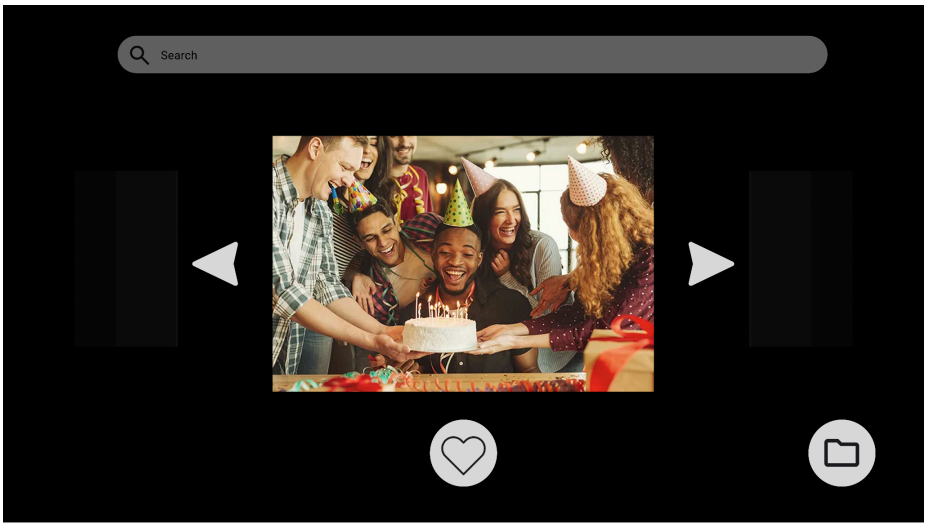


Figure 7

Signals from Layer 1 produce an initial ranked list. Users navigate left and right through this ranking and "favourite" images. "Favourited" images are stored in a "folder" such that Users can revisit relevant images without backtracking. Brain-signals, click-signals, and top-k are logged. After *"k"* images, the User is prompted to "continue". If they click "*no",* the session terminates. If they click *"yes"*, the *next "k"* results are ranked utilising signals collected in this layer.

4. View Favourites

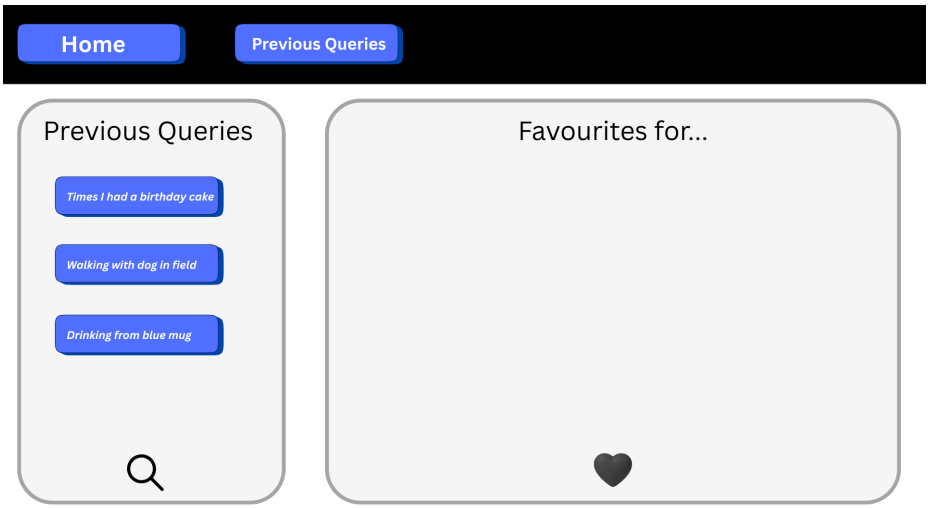


Figure 8

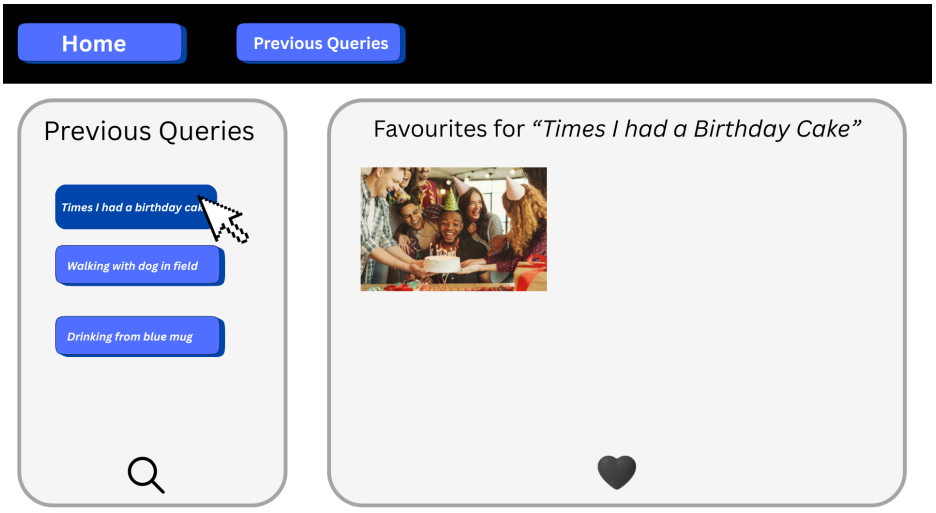


Figure 9

This exists outside the flow, and can be activated at any time (see navbar option in *Figure 5)*. The system logs "favourite'd" images such that Users can view all "favourite'd" images on a per-query basis.

3.1.1. Data Flow

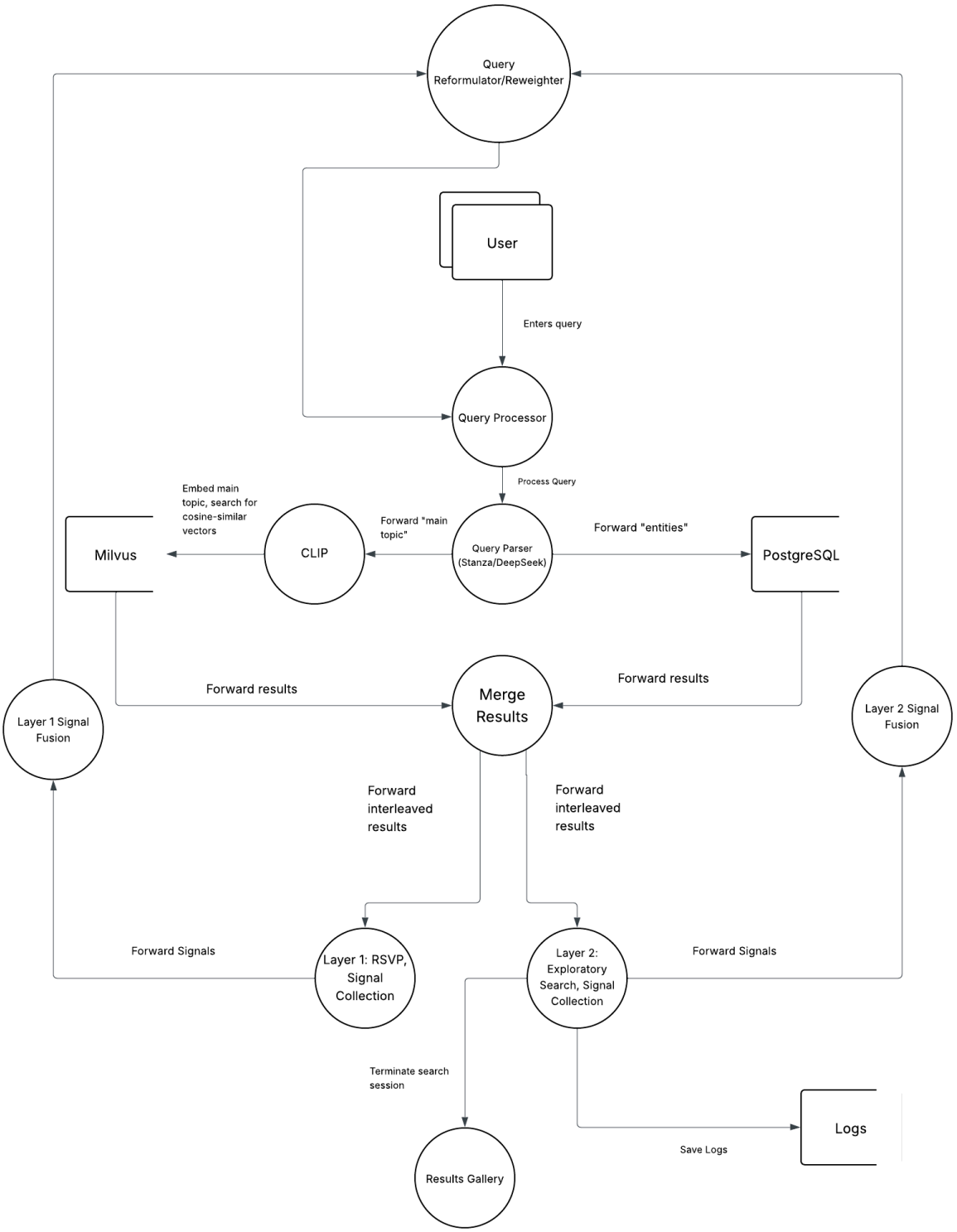
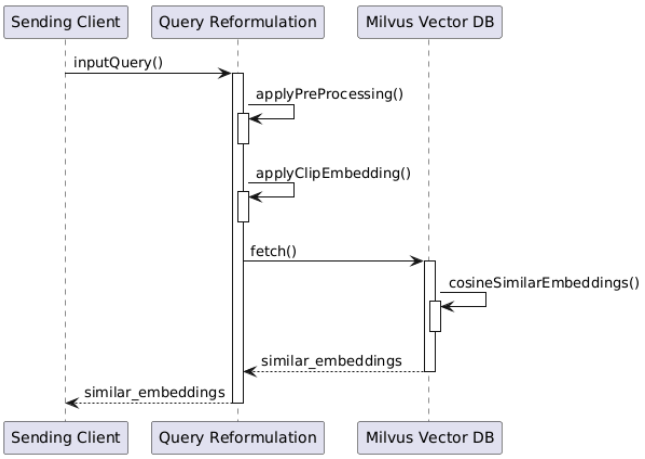


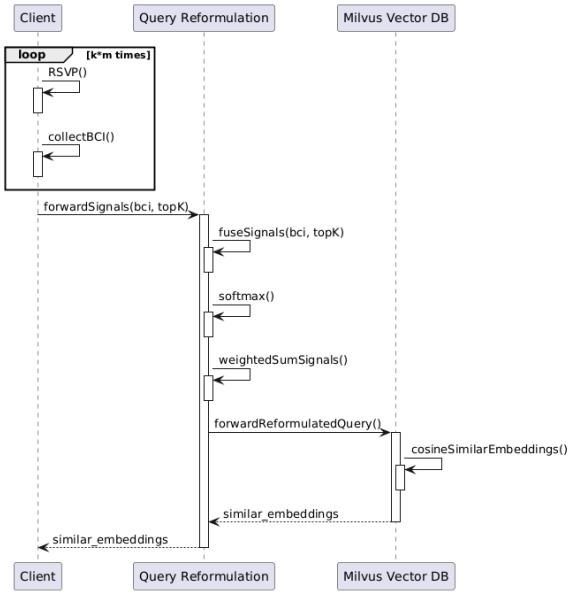
Figure 10. Data Flow Diagram

3.1.1.1. Initial Ranking



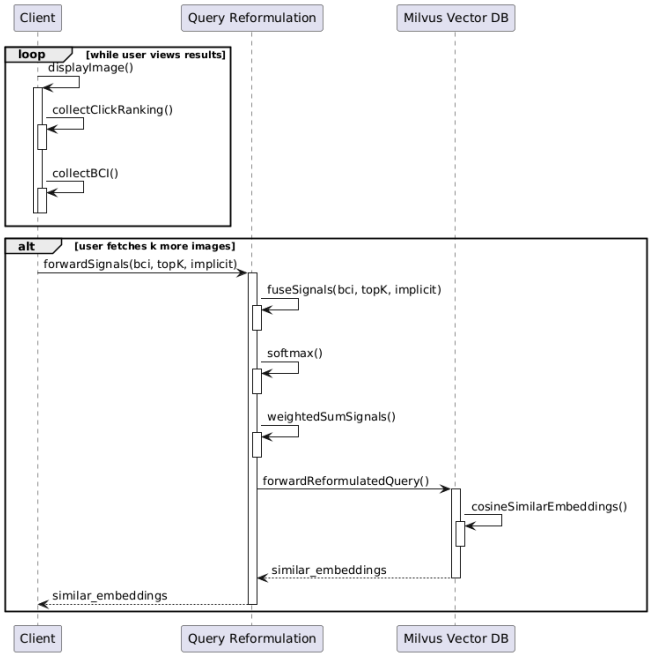
An initial ranking shall be performed by measuring the cosine-similarity of the CLIP encoding and the text query and interleaving results with "feature" matches in PostgreSQL *(Section 3.4.3*).

3.1.1.2. IRF RSVP Layer



1. First re-ranking layer *(Figure 6)*
    1.1. The system shall display one image at a time using RSVP to collect brain-signals indicating relevance, gathering implicit, rapid, brain-driven relevance-feedback akin to Healy et al. [18].
    1.2. This shall be done over *(k*m)* images, where "k" is the Top-K most relevant images, and "m" is some hyper-parameter (see *Section 4.1.4)*.
    1.3. Fusion detailed in *Section 3.2.2.3.*
    1.4. This only occurs once per query.

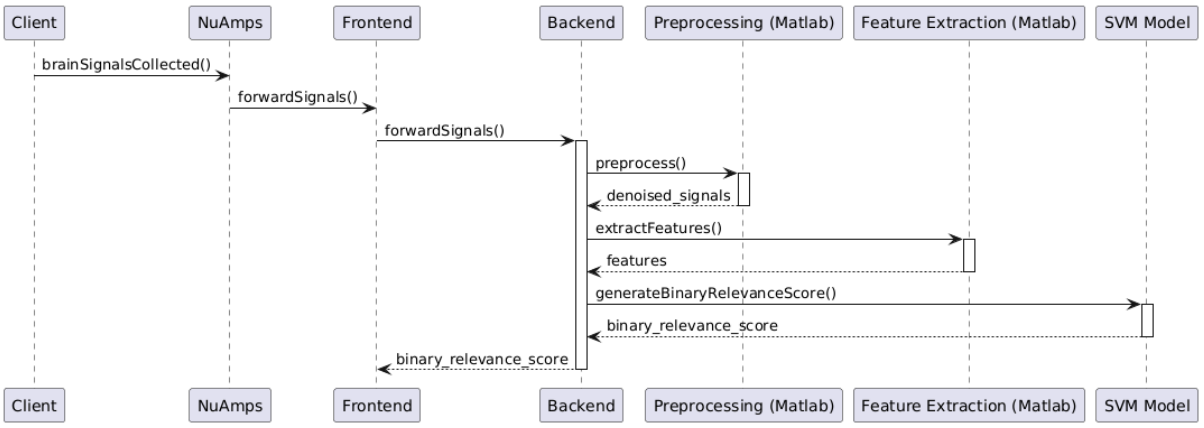
3.1.1.3. IRF Results Layer



2. Second re-ranking layer *(Figure 7*)
    2.1. Top-k images displayed one-by-one.
        2.1.1. The User shall click “next”, or “previous” to navigate.
        2.1.2. The User can “favourite” an image.
        2.1.3. The User can open a gallery of “favourite’d” images.
    2.2. While browsing, implicit-signals *(Section 4.2.2.2*) and brain-signals are collected *(Section 4.2.2.1)*.
    2.3. After *K* images, the User is prompted to “continue”.
        2.3.1. If a User “continues”, this produces *negative implicit-signals* for images *not* marked “favourite”.
        2.3.2. If a User does not seek to continue, this produces *positive* feedback (the User was able to satisfy their query).
    2.4. If the User continues searching, reweighting and reformulation will be based on signals collected thus far (implicit-feedback, brain-signals, pseudo-relevant-signals). *“Seen”* documents are removed from the candidate list.

The system alerts Users where a query has low cosine-similarity with all images (see *Heuristic Analysis, 10.1.)*.

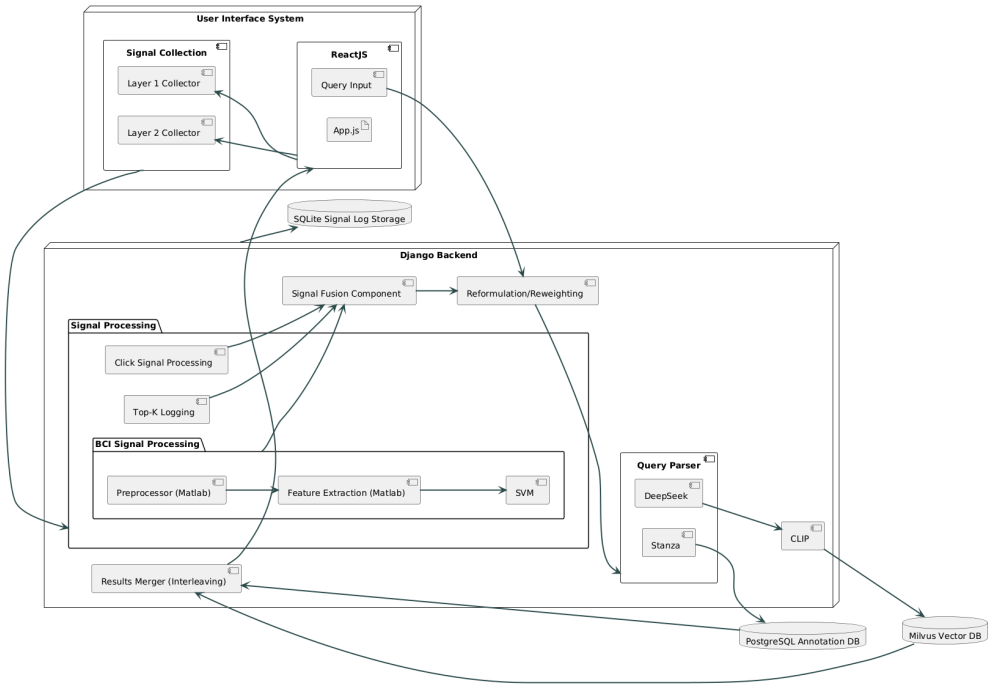
3.1.1.4. Generating Binary Scores using Brain-Signals



Brain-signals are collected and forwarded, before being preprocessed with features extracted for generating binary relevance scores *(Section 4.2.2.1)*.

3.2 System Architecture

3.2.1. System Architecture Diagram



UML Deployment Diagram

3.2.2. Component Description

3.2.2.1. ReactJS UI

3.2.2.1.1. Query Input

The user enters a text-query through the UI, where it is forwarded to the Reformulation component via the Django backend.

3.2.2.1.2. Signal Collection

3.2.2.1.2.1. Layer 1 (RSVP)

In a ReactJS UI, RSVP displays *K*M* images *(Section 3.1.1.2)* to collect brain-signals and pseudo-relevant signals.

3.2.2.1.2.1. Layer 2 (BCI, Top-K, Click)

The User is shown *K* images one-by-one in a ReactJS interface. The User views results and click-signals, brain-signals and top-k is logged *(Figure 7)*.

3.2.2.2. Signal Processing

This component collects and processes input, and forwards to Fusion. See *Section 3.3.*

3.2.2.3. Fusion

Signals are fused (see *Section 4.2.2*) and forwarded to Query Reformulation.

3.2.2.4. Query Reformulation

Initially, the query is pre-processed:

- Stemming
- Stop-word removal

We do not use “thesaurus synonyms” for query expansion, as subsequent passes will gather relevant terms for reweighting and reformulation. Reformulated queries are forwarded to Query Parser.

3.2.2.5. Query Parser

The *“Main Topic”* is extracted (using *DeepSeek*) and forwarded to CLIP. Entities are extracted using *Stanza* and searched for in PostgreSQL (mirroring Memoria [10]).

3.2.2.6. Results Merger

CLIP densely-embeds text and images into shared vector-space such that text-query-vectors can be measured against image-vectors. Textual annotations are stored in PostgreSQL. Mirroring Memoria [10], results from the “main topics” cosine-similarity search in Milvus are *interleaved* with PostgreSQL entity matches, aggregating a list of “likely” relevant images.

3.3. Algorithms

3.3.1. Signal Fusion

To extend Memoria [10], we iterate on Ye et al. [2], who fused brain-signals with pseudo-relevant signals and implicit-signals (click) to produce relevance scores on an interval (0, 1). We explain their approach as follows:

Each signal was combined in a weighted sum fusion to generate $r_i^{it,h}$ for some document h:

$$r_i^{it,h} = \theta^{it,bs} \cdot r_i^{bs} + \theta^{it,c} \cdot r_i^{c} + \theta^{it,p} \cdot r_i^{p}$$

Each signal has weighted scalars theta (bs = brain-signal, p = pseudo-relevant, c = click) and scores "r", where scores are scaled and combined to generate fused scores $r_i^{it,h}$. Superscript "it" refers to "iterative".

To expand queries, new documents are compared to *top-k ranked documents ranked by* $r_i^{it,h}$ by weighting BERT ranking using a softmax function, where documents with higher IRF scores (computed above) are favoured. This produces a feedback-enhanced score r_{h+i}^{f} for the *h+i'th* upcoming document.

$$r_{h+i}^{f} = \sum_{j=1}^{k} \frac{e^{r_{s_k}^{it,h}}}{\sum_{l=1}^{k} e^{r_{it,h}^{s_l}}} \cdot \mathrm{BERT}(d_{s_j}, d_{h+i})$$

Where k refers to top-k, document d_{sj} is in the j'th document in set S of documents with high $r_i^{it,h}$, and d_{h+i} is the h+i'th upcoming document.

Each is combined using some coefficient "c" (which balances reliance on re-ranked relevance, versus initial relevance).

$$i_t^{f} = r_{h+i} \cdot c + \mathrm{BERT}(q, d_{h+i}) \cdot (1 - c)$$

As "c" approaches 0, the original query is weighted more. Where "c" approaches 1, rf weights more.

We liken this step to the familiar [22]:

$$Q' = \alpha Q + Q_E$$

Query re-weighting algorithm presented by Jones et al. [24] for weighting expansion terms in query, where alpha scales original query with reference to Q_E.

We adapt this fusion in *Layer 2* and *Layer 1* as follows:

We replace BERT(q, d_{h+i}) in the final and second-final step with the cosine similarity of CLIP(q) and CLIP(i) (where *i* is some upcoming image). This develops a ranked-list based on user-defined relevance. Using this ranked-list, for relevant images, we index into PostgreSQL and extract "features" that are common in highly-relevant images, and use these to perform *query-expansion* and *term re-weighting.* Once this is complete, we re-send this query and retrieve a new list of results interleaved from Milvus and PostgreSQL - this time with results that align with user-defined relevant features.

Hyper-parameter selection for *c, query-expansion* and *scalars* outlined in *Section 4.1.4.*

3.4. Design Analysis and Justification

3.4.1. CLIP

CLIP embedding models feature in successful LifeLogging systems [8, 9, 10, 11]. Additionally, CLIP [6] and BERT [23] output dense embeddings, suggesting the results Ye et al. [2] achieved fusing EEG-signals in a BERT ranking *document retrieval* system could be replicated in an *image retrieval* system*,* where BERT is replaced with CLIP.

3.4.2. Multi-Layer RF using BCIs

Understanding User Interests in image-search is difficult, requiring refinement over many search cycles [24]. We select a multi-layered structure considering Nara et al. [4], who demonstrated positive results iteratively re-ranking. We integrate the work of Nara et al. and Lumina-1 [11] (RF in LifeLogging) with that of Ye et al. [2] by fusing BCI-signals with click and top-K.

Using RSVP to sort Layer 1, we extend the work of Healy et. al [18]. Explicit feedback is difficult to collect [1]; utilising RSVP to rapidly collect natural signals for RF sidesteps the problem of query ambiguity efficiently.

We note Aggarwal et al. [13], who advocate for BCI-ML utility in classification problems.

3.4.3. Cosine Similarity

Radford et al. utilise the dot product in measuring similarity (equivalent to cosine similarity where L2 normalization occurs) [6].

3.4.4. EEG and Laboratory Environment

EEG is a well-documented method of collecting brain-signals, suited particularly for lab environments. Considering this Research-and-Development project seeks *not* to deploy a system, but

to demonstrate the viability of the BCI-RF-LifeLog system, a lab-setting is appropriate to offset brain-signal variability.

We note that fMRI is demonstrably more accurate, but processing-time and extra cost [25] make it unsuitable for this system. For our context (where the system is not wholly reliant on the BCI), this precision is not worth the cost.

3.4.5. Milvus

Leading LifeLogging models [9, 10, 11, 12] utilise Milvus vector databases.

3.4.6. User-Interface

See *Heuristic Analysis* (*Appendix*, *10.1)*.

We use a red-fixation-symbol in Layer 1 and Layer 2 in line to indicate *where* an image *shall* appear [21] In Layer 2, buttons fade in after EEG input is taken. We gradually introduce interactability such that brain- activity related to physically “clicking” does not poison brain-activity related to relevance [2].

Considering our lab environment, the User Interface shall be developed using *ReactJS* to bolster portability and iterability.

3.4.7. SVM V.S. CNN for BCI Model

As noted by Liang et al. [26], CNNs outperforms SVMs in brain-decoding tasks, but require more data and time-to-process. In a fast-moving system, where brain-signals are supplemental and less-sophisticated (compared to rigorous brain-analysis tasks such as image reconstruction), the SVM is appropriate.

We utilise manual feature-extraction (FE). EEG FE is well-documented, as are the distinct ERP-effects that concern “relevance” [14]. Moreover, manual feature-extraction boasts computational efficiency advantages over automatic feature-extraction [27].

3.4.8. SVM Kernel

We consider the pervasiveness of the Gaussian function [2] to map the input “feature” vector into dimensional space, as compared to the Adaptive Gaussian (AG) function proposed by Elen et al. [28]. As such, we select the Gaussian kernel for SVM in spite of AG's noted potential (see *Section 5)*.

3.4.9. Query Parsing and Interleaving

Considering their high AD search scores, we mirror Memoria [10] in selecting *DeepSeek* and *Stanza* for query parsing. Similarly, we iterate on their interleaving of CLIP results from Milvus with PostgreSQL “entity” matches.

4. Evaluation Plan

We assume that where an IR system performs in a (valid) experimental setting, then it shall perform in operational settings [1]. Considering the variability of EEG, we conduct our evaluation in a laboratory setting.

4.1 Evaluation Design

We note the following requirements [1]:

1. Queryset
    a. LSC’25 AD and KIS queries shall be utilised [29] *(see Appendix 10.2)*.
    b. Queries are “conventional” (wherein the querier searches for relevant images), or “restricted” (where there are additional restrictions related to time, place or similar) [29].
2. A set of images labelled with a relevance-class for each test query.
    a. LSC’25 LifeLogging dataset shall be utilised [30].
    b. Images will be marked on a 4 point scale (class 1 to 4, described below) by BCI-RF-LifeLog contributors.
3. A set of Users to judge the system
    a. Users shall take the SUS [31] *(Section 4.1.2)*.
    b. Users shall match our "target-user profile”.

User Testing shall consist of:

- Structured Testing (Task 1, Unobserved)
    - A User is given a query and a set of images.
    - Users manually judge the relevance of each image given some query on a 4 point scale [1]:
        - Class 1: does the document completely satisfy the information requirement?
        - Class 2: does the document partly answer the information requirement?
        - Class 3: does the document develop the User’s understanding of the information requirement (without fulfilling it)?
        - Class 4: is the document non-relevant to the information requirement?
    - We do not observe the User in this task to minimize the Hawthorne effect [25].
- Unstructured User Testing (Task 1, Observed)
    - Users will be shown an image and must reverse engineer a query to locate that image in the dataset.

- The User is observed in this task to observe “usability”. Precautions shall be taken to reduce the Hawthorne effect [32].

4.1.1. A/B/n Testing

To evaluate the benefits of Multi-Layer RF, we perform A/B/n testing on 3 systems:

Name	Include RSVP Layer	Include Brain-Signals
S1	Y	Y
S2	N	Y
S3	N	N

We mirror Nara et. al [4] in using A/B/n to select our CLIP backbone.

- CLIP ViT-B/32 (as utilised by Gago et al. [10])
- CLIPS (as utilised by Ho-Le et. al [9] and Le-Hinh et. al [11])
- CLIP and BLIP2 (as utilised by Tran et al. [8])

We evaluate A/B/n tests using *quantitative metrics (Section 4.3.1)*.

4.1.2. Usability Testing

We utilise the System-Usability-Survey (SUS) [31] to grade Usability on a Likert Scale (where 80 or more indicates high usability).

Heuristic Analysis (see *Appendix 10.1)* must be conducted utilising Jakub Nielsen’s Usability Heuristics [33] after changing UI components.

4.1.3. A/B Testing

We conduct A/B Testing to contrast wet and dry EEG. Dry EEG is preferable (for usability, reusability, speed-of-testing), but *reliability* is of higher priority than *satisfiability* with regards to non-functional-requirements. We gauge these systems on *quantitative metrics (Section 4.3.1)*.

4.1.4. Hyper-parameter evaluation

Name	Usage	Method
K	Top-K relevant images	We evaluate an appropriate value for K in *User-Testing* through qualitative feedback.
M	In context of K*M for RSVP trials)	We evaluate an appropriate value for M in *User-Testing* through qualitative feedback, and Unit Testing.
Coefficient C	Used in reformulation post-softmax	We evaluate an appropriate value for C in automatic Unit Testing with pre-defined values and logged signal data.
Theta *bs*, Theta *c*, and Theta *p*	Weights for weighted fusion sum *(3.1.1)*.	We evaluate an appropriate value for these weights using automatic Unit Testing with pre-defined values and logged signal data.
Similarity Threshold	A minimum threshold that indicates the image is relevant to the query. This threshold is required to indicate if there are no images that are relevant to the query in the set of images.	For images marked *irrelevant* to a query, calculate the average of all cosine-similarity scores between each query and irrelevant image to develop a threshold N such that if no image within the top-K exceeds N we can say that there are no relevant images in the dataset and warn the User.

We gauge click relevance on a 4 point Likert-scale [34] calibrated through User-testing.

For query expansion, the *nature* of the features selected and the *number* of features and images we select from shall be ascertained through evaluation.

4.2 Data and Collection

4.2.1. LifeLog Dataset

We use the LSC'25 dataset (18 months of LifeLogging data) [30]. We mirror SnapSeek [9] and Memoria [10] in utilizing YOLO* and PaddleOCR for generating annotations. In keeping with Memoria, we utilize BLIP2 for captioning and Places365 for scene detection.

We iterate on Memoria in grouping by events, and by adding metadata to the Milvus Vector Database (their *Future Work)* to facilitate greater filtering.

**YOLOv11 was used by Memoria, where YOLOv8x-oiv7 and YOLOv9-E were used by SnapSeek.*

4.2.2. Collecting Signals

4.2.2.1. Brain-Signals

During *Structured User Testing* EEG-signals are collected and matched to *explicit* relevance scores the users supply until 100 signals have been collected (threshold indicated by Ye et al. [2]).

In preparing and extracting EEG-signals, we:

- Follow accepted preparatory guidelines [35]
- Preprocess to reduce noise and variability.
    - We mirror Ye et. al [2] who iterated on standard preprocessing steps while neglecting to include artifact removal techniques to improve efficiency.
- Perform feature extraction.
    - We consider Differential Entropy's comparative strength in emotion recognition tasks [36].
- Generate a binary-relevance-score from features using SVM.

When sufficient samples are collected, we perform supervised learning (as this is a "classification problem" [13]) using the SVM and gaussian kernel *(Section 3.4.8),* using scikit-learn libraries to match explicit feedback to automatic brain-signals. Signals, models and weights shall be stored in the SQLite database. Signals can also be used to perform Unit Testing to gauge appropriate values for *M (Section 4.1.4)*.

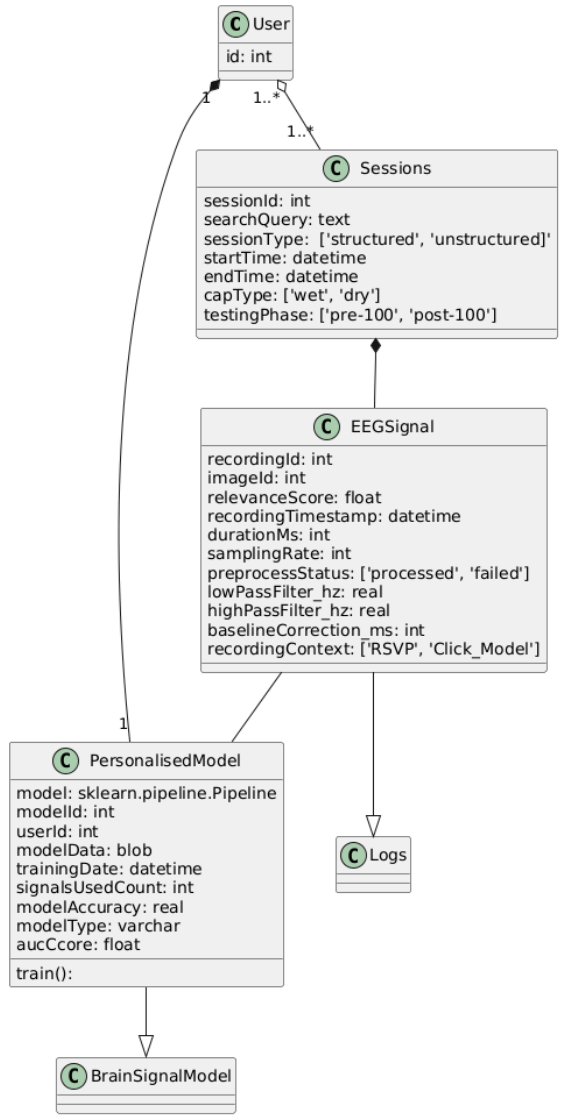


Figure 12: SqLite Database

4.2.2.2. Click-Signals and Top-K

Signals associated with a User, search session, and the set of seen documents shall be saved in a SQLite Database managed through a Django Backend.

We adapt popular click-signal collection methods (clicks to select documents from a search results list, bookmarking an item, time spent viewing item) [34] and include additional implicit-signals:

Does the:

- User navigate *back* to an image?
- User click "*continue"?*
- User *right-click* on an image?
    - *Right-clicking* opens options such as "open in new tab" or "magnify", implying interest.

We update the weight r_i^c for the i'th document based on clicks, where we gauge click-relevance with a 4 point Likert scale (totally irrelevant, semi-relevant, relevant, perfectly relevant) [2].

4.3 Criteria and metrics

4.3.1. Quantitative

To evaluate IR we use:

- Mean-Average-Precision [1]
- Precision [1]
- Recall [1]
- Average Search Time

We test the User Interface quantitatively using Nielsen's [33] Heuristic Analysis *(see Appendix: 10.1)*.

We consider DRES [37] (a platform for conducting multimedia-search evaluation used by LSC'25 [7]) thanks to its pervasiveness in multimedia evaluation.

4.3.2. Qualitative

Alongside SUS surveys [31], Users shall be surveyed for subjective feedback in regards to the utility of the system *(see Appendix, 10.3)*.

5. Future Work

There is potential to extend the BCI-RF Lifelogging system by incorporating RAG (iterating on MemoriEase [8]) to bolster performance in QA tasks, and in moving the BCI system outside of lab environments, filtering out "noise" that would exist as a result of using the system outside of a controlled environment.

Noting Quach et al. [16, 17] and Ye et al. [12], future iterations of the system could *learn* User habits to provide a system entirely controlled by BCI (word-prediction in search, SSVEP in selecting keys and options).

We recognise Liang et al. [26] and encourage the evaluation of a CNN for brain-decoding in an IRF task compared to SVM, to consider if a potentially improved accuracy is worth greater processing time and training time.

We note Ye et al. [2] in developing an RRF pipeline for RF, such that IRF data can be logged and used to improve later cycles.

We acknowledge Sluis et al. [38] in factoring eye-gaze tracking as a signal in RF. Presently, there is too much variance to say whether such a signal would positively affect the system.

Considering variability of brain-signals [13], it would be appropriate to consider the AG function [28] as the kernel for the SVM, comparing it to the Gaussian kernel.

As noted by Aggarwal et al. [13], a bottleneck in BCI-research is the calibration of a User's brain-signals. ML presents a solution for this in developing “classifiers” for signals such that systems can “guess” relevance. In developing ML models for BCI-IR wherein RF is not wholly contingent on RF, rather, where brain-signals are additional factors, we set the groundwork for developing general models in the future that would reduce the need for strict per-user calibration.

6. Conclusion

This report presents the BCI-RF LifeLogging system. The system demonstrates the potential and motivation for integrating brain-signals into RF algorithms to reduce query-ambiguity in image-search for a LifeLogging context.

We present alongside this plan a literature review and a detail of potential future work. This work presents a further step toward integrating Computational Neuroscience into Information Retrieval.

7. Group Activity Summary

Our idea came from a *CSC1103* lecture in early October, where Dr. Gareth Jones was discussing the difficulty inherent in *defining* the ASK. We had recently read a paper [37] about *decoding* brain-signals to predict relevant subsequent words. We wondered: could the problem of disambiguating the ASK be solved through the decoding of the brain-signals composing it? This naturally led us to *Relevance Feedback*, where we developed an idea for an *implicit,* unconscious, and thusly deeply *natural* derivation of the ASK.

We outline our meeting schedules:

Week 1 (06/10):
- Literature review
    - Brain-computer-interfaces
    - Relevance feedback
    - Lifelogging

Week 2 (13/10):
- Shared research
- Outline written
- Set tasks

Week 3 (20/10):
- Addressed naiveties
    - In mapping data flow, we recognised several naive assumptions.
    - We created diagrams early to identify naiveties.
- Evaluated UI.
- Set out new plan.

Week 4 (28/10):
- Draft 1 finished.
- Redrawn diagrams.
- Peer reviewed.
- Highlighted naiveties
    - No consistent target-reader
    - More precision required

Weeks 5-7 (03/11 - 17/11):
- Draft 2 finished
- Peer reviewed.

Weeks 8+ (17/11 - 28/11):
- Draft 3 finished.
- Peer reviewed.
- Iterative refining until deadline.

As is evidenced by our meetings, we took an Agile approach. We worked non-linearly; all sections were peer-reviewed by someone other than their author. We attempt to generalise these contributions below.

Contributor	Contribution
Shahin De Faoite Houshidari	Conducted literature review, developed data flow, prepared diagrams, prepared survey, prepared *related work*, responsible for sections 3, 4, 5. Responsible for editing.
Aaron Gillespie	Conducted literature review, created UI-Mockups, conducted heuristic analysis, wrote abstract and user analysis sections, editing also.
Sadhbh Murphy	Conducted literature review (lifelogging especially), developed 2.2 and 2.3, developed use case diagram. Developed data flow.
Ciaran Farrell	Conducted literature review (BCI especially). Focused on Evaluation Plan (particularly 4.2), editing document.

8. References

[1] G. Jones, “Text Retrieval: Evaluation of IR Systems,” Oct. 2025.

[2] Z. Ye et al., “Relevance Feedback with Brain Signals,” arXiv.org, Dec. 09, 2023. Available: https://arxiv.org/abs/2312.05669. [Accessed: Nov. 08, 2025]

[3] Woodberry, Emma, *et al.* “The Use of a Wearable Camera Improves Autobiographical Memory in Patients with Alzheimer’s Disease.” Memory, vol. 23, no. 3, 17 Feb. 2014, pp. 340–349, https://doi.org/10.1080/09658211.2014.886703.

[4] R. Nara et al., “Revisiting Relevance Feedback for CLIP-based Interactive Image Retrieval,” arXiv.org, 2024. Available: https://arxiv.org/abs/2404.16398. [Accessed: Nov. 08, 2025]

[5] K. Su and K. A. Robbins, “A Framework for content-based Retrieval of EEG with Applications to Neuroscience and beyond,” *PubMed*, vol. 101, pp. 1–8, Aug. 2013, doi: https://doi.org/10.1109/ijcnn.2013.6707106. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC3997173/. [Accessed: Nov. 08, 2025]

[6] A. Radford et al., “Learning Transferable Visual Models from Natural Language Supervision,” arXiv:2103.00020 [cs], Feb. 2021, Available: https://arxiv.org/abs/2103.00020. [Accessed: Nov. 08, 2025]

[7] C. Gurrin et al., “Proceedings of the 8th Annual ACM Workshop on the Lifelog Search Challenge,” LSC ’25: 8th Annual ACM Workshop on the Lifelog Search Challenge, Jun. 2025, doi: https://doi.org/10.1145/3729459

[8] Q.L. Tran, B. Nguyen, G. Jones, and C. Gurrin, “MemoriEase 3.0: A RAG-Enhanced Conversational Lifelog Retrieval System at LSC’25,” pp. 52–57, Jun. 2025, doi: https://doi.org/10.1145/3729459.3748689

[9] Minh-Quan Ho-Le, D.-K. Ho, V.-T. Ninh, T.-T. Nguyen, Cathal Gurrin, and M.-T. Tran, “SnapSeek 3.0: Interactive Lifelog Search System with Enhanced Scene Understanding,” pp. 7–14, Jun. 2025, doi: https://doi.org/10.1145/3729459.3748697

[10] A. Gago, B. Kaluza, and A. J. R. Neves, “MEMORIA: a Memory Enhancement and MOment RetrIeval Application at the LSC2025,” pp. 1–6, Jun. 2025, doi: https://doi.org/10.1145/3729459.3748693

[11] N.T. Le-Hinh , C.T. Huynh, M.Q. Ho-Le, D.K. Ho, M.T. Tran, and V.T. Huynh, “LUMINA-1: Learning and Understanding Multimedia in Immersive Navigable Archives for Lifelog Retrieval,” pp. 15–22, Jun. 2025, doi: https://doi.org/10.1145/3729459.3748698

[12] Z. Ye, Q. Ai, and Y. Liu, “Brain-Computer Interface Meets Information Retrieval: Perspective on Next-generation Information System,” BCIMM ’24: Proceedings of the 1st International Workshop on Brain-Computer Interfaces (BCI) for Multimedia Understanding, pp. 61–65, Oct. 2024, doi: https://doi.org/10.1145/3688862.3689114

[13] S. Aggarwal and N. Chugh, “Review of Machine Learning Techniques for EEG Based Brain Computer Interface,” *Archives of Computational Methods in Engineering*, vol. 29, Jan. 2022, doi: https://doi.org/10.1007/s11831-021-09684-6

[14] S. Sur and V. Sinha, “Event-related potential: an Overview,” *Industrial Psychiatry Journal*, vol. 18, no. 1, p. 70, 2009, doi: https://doi.org/10.4103/0972-6748.57865. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3016705/. [Accessed: Nov. 18, 2025]

[15] M. J. A. Eugster et al., “Natural brain-information interfaces: Recommending Information by Relevance Inferred from Human Brain Signals,” Scientific Reports, vol. 6, no. 1, Dec. 2016, doi: https://doi.org/10.1038/srep38580. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC5143956 [Accessed: Nov. 08, 2025]

[16] B. M. Quach, C. Gurrin, and G. Healy, “DERCo: a Dataset for Human Behaviour in Reading Comprehension Using EEG,” *Scientific Data*, vol. 11, no. 1, Oct. 2024, doi: https://doi.org/10.1038/s41597-024-03915-8. Available: https://www.nature.com/articles/s41597-024-03915-8. [Accessed: Nov. 22, 2025]

[17] B.M. Quach, T.B. Nguyen, C. Gurrin, and G. Healy, “Decoding EEG Signals to Explore Next-Word Predictability in the Human Brain,” *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 47, no. 0, 2025, Available: https://escholarship.org/uc/item/5jd2k0dw. [Accessed: Nov. 22, 2025]

[18] G. Healy, C. Gurrin, and A. Smeaton, “Lifelogging and EEG: Utilising Neural Signals for Sorting Lifelog Image Data,” May 2014. Available: https://doras.dcu.ie/20233/1/qs2014_poster_draft2.pdf. [Accessed: Nov. 08, 2025]

[19] C. Gurrin, *LSC’24 Task Outline*. 2024. Available: http://lifelogsearch.org/lsc/resources/LSC24-Topics-Release.pdf. [Accessed: Nov. 11, 2025]

[20] John Hopkins Medicine, “Electroencephalogram (EEG),” *John Hopkins Medicine*, 2019. Available: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electroencephalogram-eeg. [Accessed: Nov. 11, 2025]

[21] L. Thaler, A. C. Schütz, M. A. Goodale, and K. R. Gegenfurtner, “What Is the Best Fixation Target? The Effect of Target Shape on Stability of Fixational Eye Movements,” *Vision Research*, vol. 76, pp. 31–42, Jan. 2013, doi: https://doi.org/10.1016/j.visres.2012.10.012

[22] G. Jones, “Text Retrieval: Relevance Feedback,” Oct. 2025.

[23] J. Devlin, M.W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *ArXiv*, Oct. 11, 2018. Available: https://arxiv.org/abs/1810.04805. [Accessed: Nov. 22, 2025]

[24] G. Jones, “Multimedia Search,” Nov. 2025.

[25] R. Sitaram et al., “fMRI Brain-Computer Interface: a Tool for Neuroscientific Research and Treatment,” Computational Intelligence and Neuroscience, vol. 2007, pp. 1–10, Nov. 2007, doi: https://doi.org/10.1155/2007/25487. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233807/. [Accessed: Nov. 08, 2025]

[26] Y. Liang, K. Bo, S. Meyyappan, and M. Ding, “Decoding fMRI Data with Support Vector Machines and Deep Neural Networks,” Journal of Neuroscience Methods, vol. 401, p. 110004, Jan. 2024, doi: https://doi.org/10.1016/j.jneumeth.2023.110004. Available: https://www.sciencedirect.com/science/article/pii/S0165027023002236. [Accessed: Nov. 08, 2025]

[27] R. Kumar, “Feature Extraction in Machine Learning,” *Datacamp.com*, Feb. 11, 2025. Available: https://www.datacamp.com/tutorial/feature-extraction-machine-learning. [Accessed: Nov. 24, 2025]

[28] Abdullah Elen, Selçuk Baş, and Cemil Közkurt, “An Adaptive Gaussian Kernel for Support Vector Machine,” Arabian Journal for Science and Engineering, vol. 47, no. 8, pp. 10579–10588, Mar. 2022, doi: https://doi.org/10.1007/s13369-022-06654-3

[29] C. Gurrin, *LSC’25 Task Outline*. 2025. Available: https://view.officeapps.live.com/op/view.aspx?src=http%3A%2F%2Flifelogsearch.org%2Flsc%2Fresources%2FLSC2025-Topics-Release.docx&wdOrigin=BROWSELINK. [Accessed: Nov. 11, 2025]

[30] Lifelog Search Challenge. “DATA | LSC’25 at ICMR’25 – Lifelog Search Index.” *LifelogSearch.org,* Available: https://lifelogsearch.org/lsc/lsc_data/ [Accessed: Nov. 26, 2025]

[31] N. Thomas, “How to use the system usability scale (SUS) to evaluate the usability of your website,” *Usability Geek*, Jan. 15, 2019. Available: https://usabilitygeek.com/how-to-use-the-system-usability-scale-sus-to-evaluate-the-usability-of-your-website/

[32] M. Azarova, “The Hawthorne Effect or Observer Bias in User Research,” Nielsen Norman Group, May 21, 2023. Available: https://www.nngroup.com/articles/hawthorne-effect-observer-bias-user-research/

[33] J. Nielsen, “10 Heuristics for User Interface Design,” *Nielsen Norman Group*, Jan. 30, 2024. Available: https://www.nngroup.com/articles/ten-usability-heuristics/

[34] S. Jung, J. L. Herlocker, and J. Webster, “Click Data as Implicit Relevance Feedback in Web Search,” Information Processing & Management, vol. 43, no. 3, pp. 791–807, May 2007, doi: https://doi.org/10.1016/j.ipm.2006.07.021

[35] University of Galway, “Standard Operating Procedure,” 2016. Available: https://www.universityofgalway.ie/media/schoolofpsychology/files/EEG-Lab-Arts-Standard-Operating-Procedure-Version-2.3.pdf

[36] R.N. Duan, J.Y. Zhu, and B.L. Lu, “Differential Entropy Feature for EEG-based Emotion Classification,” *IEEE Xplore*, Nov. 01, 2013. doi: https://doi.org/10.1109/NER.2013.6695876. Available: https://ieeexplore.ieee.org/document/6695876. [Accessed: Nov. 24, 2025]

[37] L. Sauter, R. Gasser, H. Schuldt, A. Bernstein, and L. Rossetto, “Performance Evaluation in Multimedia Retrieval,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 21, no. 1, pp. 1–23, Dec. 2024, doi: https://doi.org/10.1145/3678881

[38] F. van der Sluis and E. L van den Broek, “Feedback beyond accuracy: Using Eye‐tracking to Detect Comprehensibility and Interest during Reading,” *Journal of the Association for Information Science and Technology*, vol. 74, no. 1, pp. 3–16, May 2022, doi: https://doi.org/10.1002/asi.24657. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10084433/. [Accessed: Nov. 10, 2025]

9. Bibliography

[1] S. Shams, R. Antonello, G. Mischler, S. Bickel, A. Mehta, and Nima Mesgarani, "Neuro2Semantic: a Transfer Learning Framework for Semantic Reconstruction of Continuous Language from Human Intracranial EEG," pp. 2920–2924, Aug. 2025, doi: https://doi.org/10.21437/interspeech.2025-1010

[2] X. Chen *et al.*, "Web Search via an Efficient and Effective Brain-Machine Interface," *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pp. 1569–1572, Feb. 2022, doi: https://doi.org/10.1145/3488560.3502185

[3] A. Ksibi, A. S. D. Alluhaidan, A. Salhi, and S. A. El-Rahman, "Overview of Lifelogging: Current Challenges and Advances," *IEEE Access*, vol. 9, pp. 62630–62641, Apr. 2021, doi: https://doi.org/10.1109/access.2021.3073469

[4] Elder , "Cathal Gurrin Talks to Elder about Lifelogging and Dementia - Insight," *Insight*, Mar. 31, 2020. Available: https://www.insight-centre.org/cathal-gurrin-talks-to-elder-about-lifelogging-and-dementia/. [Accessed: Nov. 08, 2025]

[5] S. U. Rahman, N. O'Connor, J. Lemley, and G. Healy, "An Investigation of pre-stimulus Eeg for Prediction of Driver Reaction Time," Biomedical Physics & Engineering Express, vol. 11, no. 3, pp. 035003–035003, Mar. 2025, doi: https://doi.org/10.1088/2057-1976/adbf25

[6] M. Wójcik, "Brain–computer Interface in the Context of Information Retrieval Systems in a Library," Library Hi Tech, Mar. 2021, doi: https://doi.org/10.1108/lht-09-2020-0239

[7] D. Valeriani, R. Poli, and C. Cinel, "A collaborative Brain-Computer Interface to improve human performance in a visual search task," Apr. 2015, doi: https://doi.org/10.1109/ner.2015.7146599

[8] K. M. Athukorala, "Enhancing Exploratory Information-Seeking through Interaction Modeling," Lecture Notes in Computer Science, pp. 478–483, 2014, doi: https://doi.org/10.1007/978-3-319-08786-3_43

[9] Li, B., Lin, Y., Gao, X. and Liu, Z. (2021). Enhancing the EEG classification in RSVP task by combining interval model of ERPs with spatial and temporal regions of interest. *Journal of Neural Engineering*, 18(1), p.016008. doi:https://doi.org/10.1088/1741-2552/abc8d5.

[10] J. M. Clements, E. W. Sellers, D. B. Ryan, K. Caves, L. M. Collins, and C. S. Throckmorton, "Applying Dynamic Data Collection to Improve Dry Electrode System Performance for a P300-based Brain–computer Interface," Journal of Neural Engineering, vol. 13, no. 6, pp. 066018–066018, Nov. 2016, doi: https://doi.org/10.1088/1741-2560/13/6/066018

10. Appendix

10.1. Heuristic Analysis

Heuristic	Identified Problem	Solution
Visibility of System Status	In the Begin stage, a search bar is presented for querying. If not prepared, the user may be surprised by the subsequent search process (flashing images and such).	A clickable dialogue info-box is available to indicate how the search retrieval process works.
	User not aware if system is loading or frozen	Where the system is loading, display “loading” text, and spinning circle.
	User not aware if buttons are interactable or not (In the case of skipping through images, the user must wait a pre-defined time before jumping to the next, as such the button is deactivated)	Buttons (when deactivated) are greyed or completely removed..
Match Between System and Real World	Icons like “tick” or “x” to signify “proceed” or “cancel” are too ambiguous. Cultural differences between colours/icons make their usage unreliable.	Use unambiguous English text supplemented with icons.
User Control and Freedom	If a User wants to back-track to a photo from a previous k-batch.	The system will allow for previous batch visits.
	If a User wants to unselect a photo from a far back k-batch	All selected images can be opened and accessed (similar to a folder) through an always available button, where the user can unselect.
Consistency and Standards	Users must learn what “selecting” relevant content means, an unnecessary barrier. This system should stay consistent to the user’s expectations according to other digital products.	Image selection is analogized to the “Liking” system which is used in various other external apps where media is “Liked” and placed into a “Liked” folder.
Error Prevention	User proceeds through the image wheel too quickly, skipping through possibly relevant content without enough time for RF to occur.	The user must wait a defined time (Say 2 seconds) before skipping to the next image.
	If the queried content is not in the system, the User may not find out (assuming the search system allows for endless continuation).	The system alerts the user if it is reasonably suspected that the queried content is not present.
Recognition Rather than Recall	The user must rely on their own memory to recall what images have been selected, otherwise they must backtrack	Selected images can be accessed and viewed within a “Liked” folder, which is openable through a button on screen.
Flexibility and Efficiency of Use	To officially end search, the User must explicitly click to do so after exploring k images, before querying again	The user can re-enter a query into an always available search bar which will automatically terminate the session and begin a new one.
Aesthetic and Minimalist Design	The UI must not use a visual style that would interfere with the input captured by the BCI.	The UI will feature a monochrome colour scheme, with no visual stimuli during RSVP (other than the relevant image and fixation cross).

Error Recovery	The system requires already present content, for the user to query. What if there is none?	The system alerts the user that there is no content to query, instructing them to begin lifelogging in order to build a content database.
Help and Documentation	The User may still be confused from the search process.	A more detailed, written user manual is available for the user to view.

10.2. Example Queries for System

The following queries are taken from the LSC'24 and LSC'25 specifications to give examples for the evaluation questions to be used for AD and KIS tasks.

LSC2512. Find examples of when I was ordering food in a cafe from a menu in a picture frame.



Figure 13: LSC'2512 [29]

LSC2511. Find examples of when I was paying for something with a currency that is not a EURO.

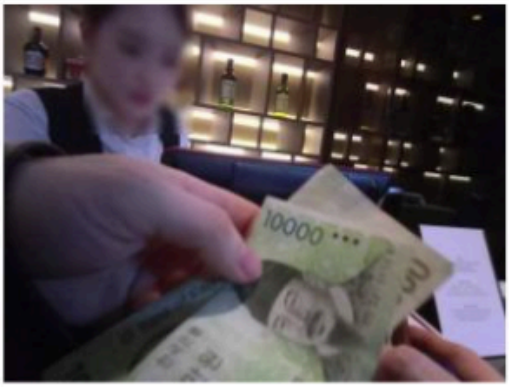


Figure 14: LSC'2511 [29]

LSC2513. Find examples of when I was drinking water in a mirror when I was not at home (i.e. not in my bathroom or kitchen).



Figure 15: LSC'2513 [29]

10.3. SUS Survey with Subjective Feedback Questions

We open with subjective feedback. Question 7 onwards contains questions from the “System Usability Survey” for evaluating Usability [31].

BCI-RF-LifeLog Search Feedback

Feedback form

* Indicates required question

1. Please select the sentiment that most closely aligns with your own in regards to the system fetching relevant images for queries? *

Mark only one oval.

- ◯ Very unreliable
- ◯ Unreliable
- ◯ Reliable
- ◯ Very reliable
- ◯ Flawless

2. Did you find the BCI to be intrusive? *

Mark only one oval.

- ◯ It was very intrusive.
- ◯ It was mildly intrusive.
- ◯ It wasn't intrusive at all.
- ◯ I have no opinion on this.

3. Did you find that (after RSVP) the first set of ranked images featured relevant images, or did you feel that you needed to continue exploring after RSVP? *

Mark only one oval.

- ◯ Very relevant images were retrieved.
- ◯ Mostly relevant images were retrieved.
- ◯ Some relevant images were retrieved.
- ◯ Few relevant images were retrieved.
- ◯ No relevant images were retrieved.

4. How did you feel about the **length of time** you spent in RSVP? *

Mark only one oval.

- ◯ Uncomfortably long.
- ◯ No thoughts on this.
- ◯ Barely noticeable.

5. Optionally, if you have any suggestions to make the feedback system more useful/the system more usable, please add them here.

6. Optionally, do you feel as though this system succeeds at de-emphasizing the need to produce precise queries? Do you feel like ambiguity is less felt when compared to your experience searching with conventional systems?

Please read the following instruction carefully.
In the following questions:

1 = Strongly **Disagree**
5 = Strongly Agree

Please tick the box that aligns most closely with your personal experience.

7. I would like to use this system frequently. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

8. I found the system unnecessarily complex. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

9. I thought the system was easy to use. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

10. I think that I would need the support of a technical person to be able to use this system. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

11. I found the various functions in this system were well integrated. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

12. I thought there was too much inconsistency in this system. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

13. I would imagine that most people would learn to use this system very quickly. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

14. I found the system very cumbersome to use. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

15. I felt very confident using the system. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

16. I needed to learn a lot of things before I could get going with this system. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

This content is neither created nor endorsed by Google.

Google Forms

Note that questions 7 and onwards are part of the SUS [31]